Deep Active Inference for State Estimation by Learning from Demonstration

Cedric De Boom, *Tim Verbelen and Bart Dhoedt. Ghent University - imec, IDLab, Department of Information Technology, Technologiepark-Zwijnaarde 15, 9052 Ghent, Belgium. E-mail: firstname.lastname@ugent.be.*

Traditionally in reinforcement learning, agents are guided towards solving a task in a given environment through a combination of reward signals [1]. For example, in the classic mountain car environment, the agent receives a large positive reward when it reaches the end goal and accumulates negative rewards for every action it performs [2, 3]. In the active inference setting, on the other hand, there is no notion of a direct reward signal; the agent is rather driven towards a particular state configuration that it favours [4]. Traditionally, these preferred states are prescribed by a prior model. Posterior state predictions based on sensor observations are then driven to match these prior states through minimization of their Kullback-Leibler (KL) divergence. At the same time, a likelihood model is trained to explain the agent's observations based on its state by optimizing the negative log-likelihood (NLL). The sum of both KL and NLL terms leads to the important concept of free energy, which can be minimized by performing actions in the environment that are dictated by a certain policy, thereby solving the task at hand.

In non-complex environments such as the mountain car, an agent's state can be defined in terms of simple physical units such as its current position and velocity. However, in more complex settings – e.g. a self-driving car or an industry-grade robot – the state of the agent cannot clearly be defined, as it has to fuse information coming from a wide range of sensors, such as LiDAR scanners, cameras, rotary encoders, etc. If we do not know the exact state upfront, encoding preferred states in the prior model is therefore not feasible.

To tackle this problem, recent advances in state-of-the-art machine learning have shown that deep neural networks are able learn rich feature representations from high-dimensional input [5]. We therefore adopt multilayer neural networks to implement the prior, posterior and likelihood models, as inspired by [6]. Variational inference and the reparameterization trick are used to calculate state distributions. Since the state is now a latent code that has to be learned during training, and for which we do not know the meaning of its individual components, it is still not possible to explicitly define a prior preference state. Therefore, in order to obtain such a prior model, we rely on expert demonstrations that provide information on the preferred states. Using this prior model, we are able to train a separate policy neural network to minimize the free energy. This way, we aim at training an agent that first reduces uncertainty on its state space by gathering observations and that at the same time exhibits goal-directed behaviour reaching states imposed by an expert. With our framework, as sketched in Figure 1, we hope to find robust policies that exhibit expert-like behaviour from a minimal amount of demonstrations.

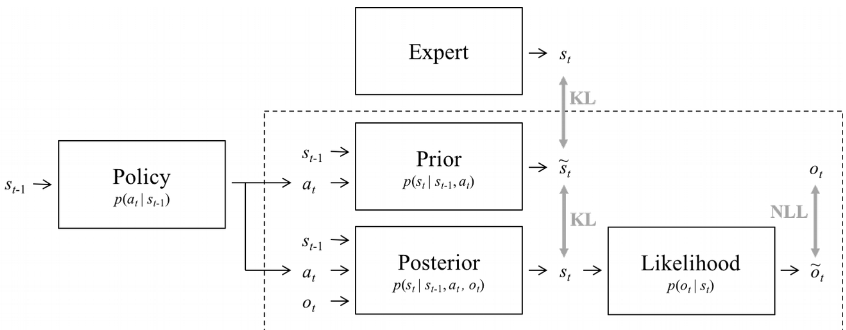

Figure 1. Graphical display of our framework. State, action and observation at time step t are written as resp. s_t, a_t and o_t.

References

[1] R. Sutton and A. Barto. *Reinforcement Learning: An Introduction*. Second Edition, MIT Press, Cambridge, 2017.

[2] A. Moore. *Efficient Memory-Based Learning for Robot Control*. PhD thesis, University of Cambridge, 1990.

[3] OpenAI Gym. *MountainCarContinuous-v0*. URL: http://gym.openai.com/envs/MountainCarContinuous-v0, 2018.

[4] K. Friston, et al. *Active Inference and Learning*. Neuroscience and Biobehavioral Reviews 68, 862-879, 2016.

[5] I. Goodfellow, et al. *Deep Learning*. MIT Press, 2016.

[6] K. Ueltzhöffer. *Deep Active Inference*. Arxiv URL: https://arxiv.org/abs/1709.02341, 2017.